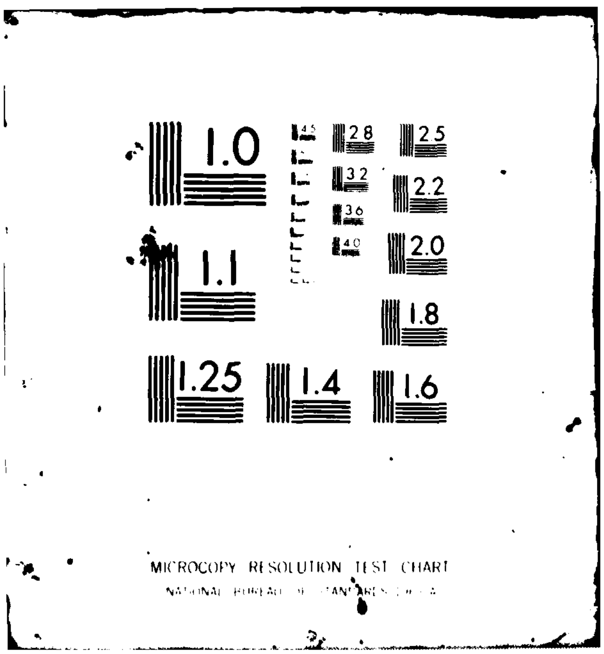
AD-A108 047 RICE UNIV HOUSTON TX DEPT OF PSYCHOLOGY F/G 5/10
TASK CHARACTERISTICS IN THE FORMATION AND USE OF UNCERTAINTY IM--ETC(U)
NOV 81 W C HOWELL N00014-78-C-0555
UNCLASSIFIED NL

END
DATE
FILMED
1 82
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

LEVEL II

(12)

RICE UNIVERSITY

Houston, Texas 77001

Department of Psychology

Research Report Series

12

TASK CHARACTERISTICS IN THE FORMATION AND USE OF UNCERTAINTY IMPRESSIONS

William C. Howell

Rice University

Final Report

November 1981

This research was supported by the Engineering Psychology Programs, Office of Naval Research, ONR Contract N00014-78-C-0555, Work Unit NR197-050.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Final Report	2. GOVT ACCESSION NO. AD-A108 047	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Task Characteristics in the Formation and Use of Uncertainty Impressions		5. TYPE OF REPORT & PERIOD COVERED 9/1/78 - 9/30/81
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) William C. Howell		8. CONTRACT OR GRANT NUMBER(s) N00014-78-C-0555
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology Rice University Houston, TX 77001		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR197-050
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE
		13. NUMBER OF PAGES 55
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

task effects, uncertainty judgment, prior expectations, effortful processing, frequency estimation, probability estimation, choice behavior, predecision processes

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(see abstract)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CONTENTS

Task Characteristics in the Formation and Use of Uncertainty Impressions: Final Report

William C. Howell

Rice University

Abstract

This report summarizes and interprets the work carried out under contract N00014-78-C-0555 over the period, 9/1/78 - 9/30/81. Fourteen studies ranging from narrowly focused molecular experiments to more broadly focused molar ones were completed on issues surrounding the general question of how people form and use impressions of event uncertainty. Particular attention was directed toward the influence of task features distinguished through an earlier taxonomic analysis. Our general hypothesis was that response demands, the defining characteristics of observed events, and prior beliefs regarding event causation play an important role in (a) how frequentistic evidence is processed, and (b) how accurately subsequent judgments or choices reflect the observed evidence.

The overall findings were consistent with this hypothesis and therefore supportive of the taxonomy, although a number of questions remain concerning the underlying cognitive processes. For example, probability estimation is consistently less accurate than frequency estimation for the same observed events, and either type of estimation

enhances the quality of subsequent choice behavior. We were able to show that the inferiority of probability to frequency estimation is not attributable to differential encoding and probably involves reliance on different kinds of stored information. However, we were not able to pinpoint what that information is. Similarly, we showed that the enhancement of choice by prior estimation is a general cuing phenomenon, but were not able to describe precisely how such cuing works. And finally, we found that prior causal beliefs do tend to influence the amount of "cognitive effort" spent in processing frequentistic events. However, the effect is mitigated by what appear to be large and consistent individual propensities, and we were not able to describe the precise interaction of task and individual influences.

The major conclusion to be drawn from this work is that rather subtle characteristics of the task setting in which observations are to be incorporated into judgments or decisions can influence how -- and how well -- people perform. Such task effects are, however, superimposed upon substantial individual differences: the two combine to determine how much "cognitive effort" a person will spend processing new information.

I. Overview of the Project

INTRODUCTION

Despite the many advances that have been made over the past two decades toward understanding and improving decision performance, a number of troublesome problems remain. Since the literature documenting this conclusion has been covered thoroughly in several recent reviews (Einhorn & Hogarth, 1981; Slovic, Fischhoff, & Lichtenstein, 1977; Hammond, McClelland, & Mumpower, 1980), the point need not be established again.

It is now fairly well accepted that unaided human decisions are subject to a variety of distortions or biases that reflect basic shortcomings in the way people process information, draw inferences from it, and eventually take action. Whether or not these distortions are "rational" in some larger sense, the practical reality is that they tend to detract from system performance. Attempts to improve the situation--whether through training ("debiasing"), aiding, or designing around the human--all require some understanding of what the limitations are, how they operate, and when they are most likely to pose serious difficulties.

Among the more significant insights that has emerged over the last few years is that task conditions have an important bearing on how people approach decision problems. Some may induce or amplify the distortions in human processing; others may promote more "optimal" strategies. That being the case, it becomes necessary to shift attention to the better understanding of decision task parameters. If, as one would hope, such tasks can

be described in terms of a manageable number of common dimensions, it shall ultimately be possible to relate these taxonomic distinctions to cognitive and behavioral consequences. That, in turn, should permit better design of decision systems.

Several attempts have been made recently to identify and classify important task features (Hammond 1980; Tversky & Kahneman, 1981). The present research was carried out within the context of one such taxonomy, a scheme concerned primarily with subjective impressions of <u>event</u> <u>uncertainty</u> (Howell & Burnett, 1988). Since the most important decision problems from a man-machine system standpoint are those in which some human judgment with respect to uncertain events or decision outcomes is required -- and it is this kind of judgment that is particularly subject to distortion -- we directed our attention to the broad question of how task parameters influence human judgment of uncertainty and uncertainty-based choice. Our working hypothesis, as reflected in the aforementioned taxonomy, was that the way decision makers (DMs) approach uncertain situations is influenced by (a) their prior understanding of causal mechanisms (i.e., what process is generating the uncertain events), (b) their expectations as to response requirements (i.e., what they will have to do with the information), and (c) the kind of events about which they are uncertain. Certain of these conditions encourage "biased" processing: others tend to promote "accuracy." The broad purpose of the research, then, was first to determine whether such task distinctions are practically meaningful, and second, to gain some insight

into the cognitive processes through which any that are formed to be meaningful operate.

Any endeavor of this kind, of course, faces several practical difficulties at the very outset. It is recognized that individuals differ widely in their approach to cognitive tasks, and further, that such differences are amplified by task complexity (since complex tasks provide more processing options, and hence more bases for differences than do simpler ones). The researcher is thus obliged to adopt a strategy aimed at either very particular task effects, which requires the use of rigorously controlled and unambiguous decision problems, or much broader ones, which requires the use of more complex and ambiguous problems. Much of our understanding of judgment and decision derives from the more rigorously structured type of problem. By contrast, the present work was carried out for the most part using the more complex, ambiguous (from DM's standpoint) type of problem. Only in this way could we hope to observe the influence of widely different task features within the same experimental context. Our major studies, therefore, were designed to examine rather molar task effects in a setting that permitted DM wide latitude in processing strategy and in which overall performance was not expected to approach the theoretical maximum. This was considered fairly representative of real-world situations that rely heavily on human judgment. As a supplement to these molar studies, a number of more limited,

more rigorously structured experiments were carried out to examine specific processing issues. Even in these "molecular" ones, however, we used tasks of realistic complexity.

The present section of this report deals with the project as a whole: its objectives, general approach, major results, major shortcomings, and overall conclusions. Our goal here is to extract all the meaning--practical and otherwise--that we can from what was done, adding where appropriate (but suitably labeled) our "informed speculations." It is intended that this overview be intelligible to the nontechnical reader, and that for the technically oriented reader, it serve as a context within which to consider the research details. Individual summaries of the major experiments will be presented in the next section.

EXPERIMENTATION

In all, 14 studies of various kinds were completed. They are perhaps best classified in terms of formality (exploratory or preliminary vs. formal experiments) and scope (molar vs. molecular as distinguished above). On this basis, there was one informal and four formal studies of the molar variety; three informal and six formal studies of the molecular variety. Several others were aborted for various reasons.

Objectives. In many real-world decision problems, DM relies upon his general impression of how likely certain events are based upon first-hand experience with those events. That experience, however, is rarely acquired in a systematic way or

with any explicit intention of formulating a subjective "statistical record." As noted above, the result is often a rather ad hoc judgment or decision, which we know is subject to certain biases. However, we know very little about how people form these general impressions--how much of the available information they actually absorb--and what role certain task features play in that process.

The main objective of the project, therefore, was to determine whether the task features identified previously do indeed affect the formation and use of uncertainty impressions for "naturally occurring" events. A secondary objective was to learn as much as possible about the processes involved in any such effects recognizing, however, that our complex task requirement places severe limitations on the ability to isolate specific processes.

The essential logic of the approach was to create a situation in which a variety of attended-to events occurred naturally over time in a repetitive but nonsystematic fashion. Subjects participated in this ongoing activity, thereby having the opportunity to acquire experience with the events. Subsequently their (impressions) of event certainty was probed using both judgment and decision criteria. In view of this logic, plus the fact that there were certain task parameters that were to be varied, a great deal of attention was devoted to the development of appropriate task scenarios.

Tasks. Several task requirements follow from the above objectives.

To provide the necessary conditions, an experimental "vehicle" must: (a) generate a variety of identifiable events (b) that can be programmed to occur with varying frequencies over time (c) within a "scenario" that appears natural and meaningful to the subject, and (d) forces him to give some attention to each event occurrence. Two task scenarios were developed with these general features in mind. One, which was used primarily to study acquisition of intuitive "frequency records," was modelled after the personnel-selection problem: repetitive "events" were applicants with certain specified characteristics. The other, which was used in the large-scale studies of task parameters, was a simulated resource allocation system: "events" were specific kinds of occurrence in the environment or responses to those occurrences.

1. Personnel-selection problem. The subject was placed in the role of a personnel decision maker who evaluated (or selected) candidates for specified positions on the basis of a standard set of characteristics (e.g. sex, age, aptitudes etc.). The advertized purpose of the studies was to infer ("capture") individuals' policies from their decision behavior using statistical techniques (regression analyses). While this was, in fact, done in most of the studies, the principal goal was to explore the process by which impressions of frequency for various event features built up or shifted. This was indexed for the most part, by direct frequency estimation.

2. Resource-allocation problem. Here, the model was an integrated emergency service system for a hypothetical city. The DM functioned as an allocator and dispatcher of specific services (e.g. ambulance, police) to specific locations (e.g. city zones) in response to programmed emergency "calls." An additional decision element was involved in that calls could be verified, at some cost, prior to a dispatching response since in some areas the frequency of false alarms was (realistically) high. In essence, then, DM served in this capacity long enough to become "experienced" in the emergency tendencies, and in the effectiveness of various response strategies to those tendencies, for the entire system. And he acquired the knowledge in a fairly natural way. The whole scenario was programmed on a TRS-80 microcomputer, a fact which undoubtedly contributed to the high level of involvement and interest reported by the subjects.

Types of studies and major results. Three series of experiments were carried out--two of the molecular variety and one of the molar type.

The first series of molecular studies used the personnel selection problem, in this case as applied to evaluation of applicants for college admission. The subjects rated a large number of profiles under instructions to strive for consistency in whatever policy they chose to adopt. Profiles consisted of standard application-blank information including academic credentials, race, and sex. After rating a prescribed number of

profiles, the subject (DM) was required to give frequency or proportion estimates for race, sex, or other designated applicant features.

The initial studies in this series extended some ideas from other areas of cognitive research to the frequency-processing domain. In particular, we explored some implications of the currently popular view that personal theories of causation guide the way people process further evidence: people tend to seek out or emphasize confirming instances, and to ignore or rationalize disconfirming instances (Ross & Anderson, 1980; Ross & Lepper, 1981; Snyder & Swann, 1978). In the present context, this view would suggest that how intensively DM processes each observed event (applicant profile) should depend upon his prior beliefs regarding their causation. Thus his impression of the frequency with which specific events had occurred would be shaped by his <u>prior generator beliefs</u>. Carrying the idea a step farther, we might expect disconfirming evidence (i.e. observed frequencies inconsistent with his prior generator beliefs) to affect his posterior impression in a way quite different from that posited by most opinion-revision models. In the latter view, change is gradual and systematic: An original opinion is eroded by each successive disconfirming instance, and is gradually replaced by a new causal theory. By contrast, we are suggesting that prior beliefs may render disconfirming observations quite ineffective; that until something happens to "shake" those beliefs, observed event frequency will have little effect upon frequency impressions or judgments.

The above conception of how prior beliefs influence event frequency perception leads to several hypotheses that, in one way or another, were addressed in our first series of studies. One is simply that changes in the event generator are likely not to be detected at all, or if detected, are likely to produce rapid rather than gradual shifts in frequency estimation. Another is that certain task features such as cuing, shift magnitude, and salience of events, might play an important role in the extent to which prior beliefs control the processing of evidence. That is, such features might determine whether DM is inclined to question his prior "theory" and, therefore, look to the "evidence" in search of an alternative.

The five studies directed toward these questions generally took the following form: (a) prior beliefs were established and verified, (b) a shift in the frequency generator was introduced, and (c) frequency estimates were obtained and plotted as a function of accumulated evidence from the new generator. While far from conclusive, the results generally supported our prediction that frequency estimates are more likely to shift abruptly than gradually. However, the most noteworthy feature of the data was the magnitude of individual differences in response to the altered generator: some subjects "picked up" the change very quickly and adjusted their estimates accordingly; others relied exclusively on their prior beliefs (all but ignoring the evidence); still others shifted rapidly after a large number of observations. Very

few showed any evidence of a gradual shift. Another noteworthy finding was that it took rather blatant task-related cues to override these personal tendencies. Our conclusion, therefore, was that people differ considerably in their tendency to process disconfirming evidence of event frequency, and that these tendencies are fairly well established.

The second series of experiments grew out of this first one, and had to do with the manner in which task or individual-difference variables might operate to influence the formation (or shift) of event frequency impressions. It has been suggested elsewhere that people encode the frequency of event occurrences in a rather automatic fashion; that no special "set" is required (Hasher & Chromiak, 1977; Hasher & Zacks, 1979). Based on the results of our first studies, however, we suspected that "automatic processing" may be unique to certain laboratory tasks and not a typical real-world phenomenon. We postulated that the impression of frequency depends upon a person's investment of "cognitive effort" in encoding that specific attribute of observed events. While this investment may occur without special cuing in the case of simple memory tasks (e.g. word lists), and therefore seem not to draw upon the individual's limited-processing capacity, it would be expected to show up in a more realistically demanding task (such as selection).

The four studies in this series, therefore, addressed the question of whether special cuing influences frequency estimation

in a version of our selection task, and, if it does, whether it detracts from other processing that is taking place at the same time. If frequency is encoded automatically, it should be insensitive to specific cuing; if it requires no "cognitive effort," it should not detract from performance of a competing task. Separate experiments were designed to answer these two aspects of the "automatic endoding" issue. The first required subjects to judge and sort profiles into accept and reject categories with instructions to pay particular attention to the frequency of rejected applicants in "protected" categories. Subsequently, frequency estimates were obtained for designated types of applicants in both the reject (cued) and accept (non-cued) categories. The results clearly favored the cued category, thereby supporting our contention that frequency encoding is not entirely automatic.

The other major experiment varied the emphasis placed upon the frequency-encoding and applicant-evaluation aspects of the task in a dual-task paradigm. If no special "cognitive effort" is required to encode frequency, there should be no decrement in evaluation performance when subjects are cued for frequency estimation. As it turned out, there was a direct performance trade-off, suggesting that frequency encoding does reduce the subject's available "processing capacity." Our conclusion, then, was that in both sensitivity-to-cuing and demand-on-capacity respects, frequency encoding is not an automatic process.

The third series of studies consisted entirely of molar-level experiments using the resource-allocation problem. These studies

required highly practiced subjects and, as a result, extended over rather long periods of time. The first two were designed to explore several of the task distinctions from our taxonomy within the same basic scenario. Specifically, we were interested in whether (a) different response requirements or ways of indexing uncertainty yield comparable values for the same frequentistic events, (b) different kinds of repeated events yield estimates of comparable quality, and (c) decision or choice behavior is influenced by prior judgments of uncertainty.

The design used in these studies consisted of three groups of subjects, all of whom processed some 600 emergency calls over five daily "shifts." During the final shift, all were required to make decisions among specified event pairs purely on the basis of which one they considered more likely to occur next. Two of the groups, however, also made estimates of event uncertainty prior to the choice task: for one, the _FE_ group, it was a numerical estimate of specified event frequencies; for the other, the _PE_ group, it was a numerical probability estimate. Events about which they were queried included a variety of spatial, nonspatial, and own-performance items covering a wide range of actual frequency values. The third group, _C_, was a decision-only control. If response requirement is an important task characteristic, as we hypothesized, then _PE_ and _FE_ should produce different estimates of the same events; if event characteristics are important, then probed items categories should have an influence on estimation; and if prior estimation affect decisions, then both _PE_ and _FE_ groups should make better

decisions than the C group.

All three of these predicted effects were obtained in both studies. The estimation of event frequency was consistently more accurate than that of probability, and this difference (as well as the absolute level of accuracy) was consistently better for some types of items than for others. We were not able to establish definitively the specific aspects of events that are conducive or nonconducive to accurate processing. There was a suggestion in the data that our taxonomic distinction between externally generated and partially self-generated events may be one important feature, but the complexity of the overall task prevented a direct test of this possibility. We strongly suspect, however, that event characteristics play an important role in the way people encode observations and hence commit them initially to "storage." People might, for example, have a predisposition to "tabulate" occurrences by spatial reference more easily than by certain other qualitative or quantitative features. This issue has considerable theoretical as well as practical significance and should be addressed more directly in future research. It was a peripheral consideration in the present work.

The FE-PE difference, on the other hand, was of central interest, and is clearly not attributable to differential encoding: the groups were not aware of the estimation required until their impressions had already been formed. Therefore, the important manner in which response requirement influences estimation performance must involve the use of different stored information.

Frequency estimation produces a fairly unadulterated impression of what happened; probability estimation includes other considerations which, in a relatively stable environment, can only detract from performance (e.g., "playing hunches," emphasizing recent or vivid cases, etc.).

Finally, the fact that both estimation groups made better decisions than the non-estimation control was congruent with our expectations and was also a fundamental consideration in shaping the other studies in the series. The question was, "Why does estimation improve the quality of subsequent decisions?" Is it because of a general cuing process which directs DM's attention to certain kinds of stored information when he faces a choice, or is it because of the fact that an estimate gives him a handy summary code which he can apply directly to the decision? In the latter case, one would expect the quality of decisions to reflect the quality of estimation; in the former (general cuing) case, estimate quality would be less critical. The results of the first studies were equivocal on this point: the FE group (which made better estimates) did make slightly better choices than the PE group on the average, but not significantly so. Therefore, in the last two studies, we attempted to manipulate more directly the initial encoding and subsequent cuing processes.

In the first, PE and FE subjects were instructed at the outset that they would eventually be required to make estimates; but in half of each group the instructions were congruent with the actual estimation requirements (FE-FE or PE-PE), while in the other half

they were incongruent (_FE-PE_ or _PE-FE_). If general cuing is the main process controlling decision quality, then all of these groups should perform better on the subsequent choices than a non-cued _C_ group. If congruency of cuing affects quality of estimates, however, and the estimate is applied directly to the decision, then congruent groups (especially _FE-FE_) should do better than incongruent groups on the choice trials. As it turned out, estimation groups (all of which were now cued from the outset) again made better decisions than th control group, and by about the same margin as in the previous studies. Moreover, congruent cuing produced generally superior _choices_. Quite unexpectedly, however, it did not produce superior _estimates_. Rather, as in the earlier studies, the actual response requirement was the controlling factor: both _FE_ groups made more accurate estimates than did the _PE_ groups, and by about the same amount. Therefore, the _quality_ of estimates per se does not seem as important as _some_ form of cuing, a finding that supports the general cuing hypothesis. However, incongruency in cuing does appear to detract from decision performance, but not because of poorer "intuitive records."

The last study in this series examined several cuing issues within the context of conclusions reached in the course of the earlier "molecular" studies. The first question again concerned the general cuing hypothesis: if prior estimation improves decisions through general cuing, then would not the same result occur if the cuing were provided directly via instructions? Thus two groups were compared on decision accuracy, a _prior estimation_ group (as in

the previous studies) and a direct cuing group (that was simply told to rely on observed frequencies in making choices). The second question involved the limits of the general cuing hypotheses: if estimation or direct cuing encourages the DM to rely on observed frequencies to a greater extent than he would otherwise, then what happens if (as in our "molecular" studies) these frequencies are inconsistent with prior generator beliefs? Thus a second variable, consistency vs. inconsistency of evidence with prior set, was introduced within subjects.

The general task was in most respects identical to the others in the resource-allocation series. However, the estimation group gave only frequency estimates prior to the decision trials whereas the non-estimation (in this case, the direct cuing) group was instructed at that time to rely on frequency differences in making choices. The general cuing hypothesis would predict little, if any, difference between choices produced by the two groups, but a considerable superiority of consistent over inconsistent cuing in both groups. To the extent that DM can recognize (and adjust his "intuitive frequency record" in accordance with) the inconsistencies, one would expect both estimation and decision performance to improve. Between-group differences, of course, or substantial interactions with the consistency variable would throw the general cuing hypothesis into serious question.

The results were entirely congruent with the general cuing hypothesis: both types of cuing produced similar decision performance, and inconsistent cuing reduced the accuracy of both

groups significantly. Although we did not include a non-cued control group in this study, the inconsistent cuing performance was at about the same level as the non-cued controls in our other studies. Therefore, it would appear that prior generator beliefs bias the "intuitive frequency records" of at least some DMs (as we saw in the molecular studies), and thereby reduce the mean accuracy of choices to a level close to that produced in the total absence of cuing.

CONCLUSIONS

Viewed in the broadest terms, these studies offer considerable support for the proposition that the impressions people form of event uncertainty and the decisions they make on the basis of that uncertainty are importantly affected by identifiable task features. In this respect, they also support certain of the taxonomic distinctions and ideas proposed by Howell and Burnett (1978).

Particularly noteworthy with regard to estimation performance were (a) the consistent superiority of the frequency over the probability response mode, (b) the failure of prior set to alter this difference, but (c) the demonstration that prior set, competing tasks, and individual propensities can affect the quality of event frequency judgments. These, together with other findings, suggest somewhat the following sequence of influences leading to an uncertainty judgment. The individual's propensity to rely on prior

beliefs vs. process new evidence combines with the cuing afforded by the task to determine how much attentive effort he invests in subsequent observation of salient events. In general, more effort pays off in a better intuitive "frequency record" (the term "record," however, should not be taken too literally since, as noted, it probably exists in somewhat different forms for different people). However, the judgments produced at this point reflect the specific form of the response demand: a frequency probe will produce the best indication of whatever "record" the person has stored, whereas a probability probe will yield an estimate based at least in part on other considerations.

From a practical standpoint, actual choice behavior is more critical than judgment. In this respect, our results suggest that estimation improves subsequent decision-making. Since the amount of the improvement is about the same as that obtained by directly instructing people to consider observed frequencies, it would appear that the underlying mechanism is general cuing. This does not mean, of course, that the quality of decisions is totally insensitive to the quality of estimation: cuing a very poor impression of observed evidence could hardly be expected to aid performance, and it appears not to. The point is, however, that estimation encourages the DM to use his impression of what he has observed when making a decision, and this impression will usually be better than whatever he uses otherwise.

Given the complexity of the cognitive processes involved and

of the realistic task settings in which we studied them, it was to be expected that many questions would remain wholly or partially unanswered. The way people process uncertainty does appear to depend, at least in part, upon the kind of events that are uncertain and the complexity of the setting in which they occur. Highly salient or attention-demanding ones would seem to have an edge over less pronounced ones in the critical encoding and storage processes. While we were able to show that estimation performance is sensitive to event type, however, we made little progress toward pinning down the specific event features that are of greatest importance. There was some evidence consistent with the proposition that externally generated events produce more accurate impressions than those which are partly under the observer's control, and that spatiality may enhance event frequency encoding, but the task scenarios used did not permit rigorous test of either hypothesis.

Another suggestive--but not definitive--finding with respect to event properties was that salience per se does not guarantee establishment of a good "intuitive record" for observed frequencies. One may be able to "attend" to an event, but direct more or less of that "attentive effort" toward forming an impression of its uncertainty. This possibility, too, deserves further investigation under more rigorously controlled conditions. It is inconsistent with a popular view that holds frequency processing to be a largely "automatic" by-product of encoding.

Finally, the project did not address at all the distinction

between frequentistic and non-frequentistic events or that involving the span of events over which a judgment of future occurrences (e.g. a probability estimate) applies. The Howell and Burnett taxonomy suggested that both should play a part in the cognitive approach adopted by the DM; both should therefore influence his judgment or decision output. These issues could be studies within the present task context. The fact that they were neglected was simply a matter of available time and priorities.

PRACTICAL IMPLICATIONS

To the extent that the present project has advanced our understanding of the way in which task features contribute to human impressions and use of uncertainty information, it should be of general value in the design or evaluation of decision systems. The reader, of course, is in a better position than the authors to judge the ways in which such applications would be most useful. It might be well, however, to illustrate a few kinds of potentially useful implications.

1. Human capability. To acquire and maintain even a rudimentary impression of the status of all the events in our complex resource-allocation problem is no mean feat. That subjects were able to choose the better alternative on around 70% of the decisions posed after no more than five hours of observations speaks well for human capability in forming such impressions. However, this seems to be a capability that is subject to considerable individual differences: some individuals operate well above chance, are sensitive to shifts in generator properties, seem to weigh observed

evidence carefully; others operate near chance, rely heavily on prior beliefs, seem to all but ignore observed evidence. Therefore,

a. man is far from helpless when forced to rely on "intuition" in frequentistic decision settings, but

b. where selection is possible, this may prove a strategy worth developing toward improvement of system performance.

2. Task design. Despite individual differences, task features can produce overall improvement in the quality of "intuitive records." They can also affect the quality of uncertainty-based choices. In particular,

a. instructions or other task features that emphasize the importance of observing and monitoring specific frequentistic events can result in better intuitive "record keeping",

b. requiring estimates prior to decisions may enhance the quality of individual decisions,

c. if human estimates are to be used for purposes other than individual choice (e.g. as a basis for a group decision or an algorithmic solution), then a frequency mode will produce more veridical estimates than a probability mode.

3. Training possibilities. The discovery that people often rely on "heuristic processes" in making decisions, and that systematic errors result, led to the suggestion that DM might be debiased through training (Fischhoff, 1981). The approach has so far not

proven too successful, a finding that Hammond (1980) considers consistent with his "cognitive continuum theory": "heuristics" represent an intuitive mode of thought which, unlike analytic modes, are not conducive to subjective modification. Viewed in this context, the present results suggest that the formation of frequency impressions is, to an extent, subject to modification: thus it may lie at some intermediate point on the "cognitive continuum." Therefore,

a. it may be that individuals who show a limited propensity to process frequentistic evidence can be trained in the strategic aspects of this task, however

b. the specific nature of "trainable" and "untrainable" skills remain to to be determined, but

c. conditions specified in Hammond's taxonomy (such as time and complexity limitations) probably set the practical upper boundary.

II. Summary of Major Experiments

The full details of the 14 experiments can be found in the quarterly progress report and in the publications cited in the last section. The purpose of this section is simply to provide a summary of the designs and results of the most representative studies.

FREQUENCY ESTIMATION SERIES

1. Frequency estimation as a function of prior beliefs and amount of discrepant evidence. The purpose was to determine how

estimates shift from initially held "generator beliefs" to new values on the basis of observed frequency evidence that is inconsistent with those beliefs. One hypothesis was that prior beliefs influence the actual processing of new evidence: that small-to-moderate discrepancies may be totally ignored, but that larger ones may induce abrupt shifts in estimation toward the new values.

A between-groups design was used in which amount of discrepancy (3 levels) was crossed with number of observations (4 levels). The 120 subjects were assigned randomly to the 12 groups, and each performed a number of trials on our personnel-selection problem for hypothetical college applicants. The characteristics of those applicants differed systematically from the subject's (measured) prior expectations on key traits. Finally, he or she gave frequency estimates for those observed traits.

The results are shown in Table 1. The shift magnitude effect

Table 1 about here

----

was highly significant, $F(2,108) = 18.80$, $p = .000$, whereas the evidence effect was not, $F(3,108) = 1.15$, $p = .330$. The interaction approached but did not reach significance, $F(6,108) = 1.94$, $p = .080$. These findings suggest that prior beliefs have a powerful influence on the processing of evidence, and although observations are reflected to some degree in frequency estimates, there is no systematic trend (as most opinion revision models would predict). However, individual differences were large, and the between-subjects design was not

appropriate for examining individual functions. Therefore, the next study explored individual shift patterns.

2. Individual functions in estimation based on prior belief-evidence discrepancies. The same general approach was used as in #1 above except that subjects made a succession of frequency estimates as evidence accumulated. Thus it was possible to plot individual functions over four evidence blocks for the 20 subjects. Only one discrepancy level was used: the value (20%) that produced a reliable effect in the previous study.

The principal finding was that individuals show marked differences in apparent tendency to rely on prior beliefs or to incorporate observations in their successive frequency estimates. In a qualitative sense, virtually none of the functions showed the kind of gradual shift that would be expected if people revised prior opinion according to the typical revision models. A post-hoc cluster analysis suggested that the subjects could be grouped into two main categories: those who abandoned their prior beliefs quickly and attempted to form a new hypothesis based on their observations; and those who responded to the evidence, but settled on a compromise between their prior beliefs and their observations. Mean functions for these two "clusters" are contrasted in Figure 1.

Figure 1 about here

While the individual functions within each cluster varied somewhat in precise level and shape, the general form of the distinction

shown in the figure (i.e. shift to a new level at or below the observed frequencies for System 2; shift to an intermediate level in System 1) was quite representative of the respective groups.

While it is always dangerous to take post-hoc analyses too seriously--and it is quite unlikely that all DM's would fall into one of these two categories--we cannot escape the conclusion that individuals have widely differing propensities for either "sticking with" their initial impressions of event frequency or processing discrepant evidence.

ATTENTION ALLOCATION SERIES

The studies on frequency estimation suggested that for some subjects, prior beliefs are eroded very slowly, if at all, by contradictory evidence, while for others the shift to new (evidence) values is rapid and complete. Task characteristics that emphasize the nonstationarity of the event generating process can induce better "tracking" of the evidence, but do not seem as powerful as the individual propensities.

These and other findings implied that the formation and revision of frequency impressions is not an "automatic" by-product of dealing with the individual events, but--like other memory processes--requires special attentive effort. The studies in this series tested various aspects of this "attention-allocation hypothesis."

1. A study of event importance in frequency estimation. This study asked the question whether characteristics that individuals regard as particularly important when viewing event occurrences

produce better frequency impressions than those they consider less critical. If so, it would suggest that the amount of sheer attention accorded events is what controls the frequency impression; if not, one would suspect that attention must be directed explicitly toward the frequency attribute for it to influence judgment.

A version of the personnel-selection task was presented to four pre-selected subjects,* following which they were required to estimate frequencies for each of the seven evaluative discussions. Using the policy-capturing approach, raw-score regression weights were computed for each dimension to determine the importance attached by each individul to that predictive feature. Assuming that subjects are prone to invest greater "attentive effort" in those features they deem important, we hypothesized that frequency estimation proficiency should correspond to the cue-utilization weights if sheer attention is the controlling factor.

The results showed absolutely no relationship between the regression weights and quality of frequency judgments for any of the subjects. Thus we concluded that sheer attention to a repetitive event (in this case, an evaluative feature) is not sufficient to control the formation of a frequency impression. Either attention is irrelevant, which seemed unlikely given our earlier findings, or it must be invested more specifically in frequency encoding processes.

---

*Selected on the basis of prior evidence of their ability to control their attention to various inputs.

The latter possibility formed the basis for the next two studies.

2. Cued vs. non-cued estimation. The purpose of this experiment was to determine whether specific cuing for a subsequent frequency judgment improves the quality of that estimate. The "automatic encoding" view holds that such cuing should be ineffective, whereas the "effortful processing" position suggested by our earlier studies holds that it should.

Two groups of 12 subjects each again performed the personnel selection task, one under instructions to pay particular attention to the frequency of rejected females; and the other under similar instructions for rejected minorities. After processing 80 "applicants," all 24 subjects made estimates of the number of females or minorities rejected (i.e. the cued events). Immediately thereafter, they made a similar estimate for the accepted applicants (i.e. the non-cued events). Hence the design involved a between-group comparison of cued event type (sex vs. race) and a within-group comparison of the key cuing variable (reject vs. accept categories).

Results, as expected, showed no between-group effect $F(1,22) = 1.07$, $p = .312$; however, cuing improved estimation in both groups, $F(1,22) = 9.00$, $p = .007$. Thus it would appear that special processing invested in the frequency attribute can produce superior performance. This is consistent with the "effortful processing" hypothesis.

3. Frequency processing in a dual-task paradigm. The issue here was whether the special processing required in forming a

frequency impression occurs at the expense of other ongoing cognitive activities. Since it is generally assumed that man's "processing capacity" is limited, one might expect the greater "effort" induced by frequency cuing to detract from performance of a simultaneous task--in this case, the evaluation and choice of applicants. The "automatic processing" hypothesis, by contrast, would predict no decrement in the selection task as a function of adding the frequency requirement.

Very simply, a measure of the subject's consistency in rating applicants was devised using a policy-capturing strategy. Four subjects were required to perform 200 profile evaluations under two sets of instructions: one in which frequency estimation was emphasized (with rating secondary); the other in which rating was primary (with frequency estimation secondary). If the two demands draw upon a common "processing capacity," emphasis of either should reduce performance on the competing task.

The results were consistent with this expectation as shown in Tables 2 and 3. The superiority of frequency estimation (lower

----

Table 2, 3 about here

----

error scores) obtained under primary vs. secondary frequency instructions (Table 2) was significant, $F(1,3) = 11.363$, $p < .043$; the reliability of the differences in rating performance (Table 3) was even more so, $\chi^2(8) = 49.757$, $p < .001$. Apparently, therefore, improvement in one task as a function of instructional emphasis is

is achieved only at the expense of the other. An individual forms an impression of frequency, it would seem, by drawing upon his limited "processing capacity."

COMPLEX SIMULATION SERIES

The broad purpose of this series (which, incidentally, constituted the main business of the project) was to determine whether taxonomic distinctions involving response modes and event types produce different uncertainty measures in a realistically complex task setting. Since, as noted earlier, the initial results were quite positive, the later studies focused particularly on the relationship of estimation to choice.

1. Response mode and event type effects. The complex resource-allocation problem was used, with the distribution of emergency calls shown in Table 4 presented in a randomized order over each

---

Table 4 about here

---

of five dispatching sessions. Subjects were required to make one of several alternative responses to each call, and a record of dispatching performance was maintained and displayed on the CRT together with a map of the hypothetical city.

At the conclusion of the fourth session, subjects in two estimation groups gave frequency (FE) or probability (PE) judgments in response to "probes" of the sort illustrated in Table 5. During

---

Table 5 about here

---

the fifth session, dispatching trials were interleaved with predictive choice trials (i.e. "pick a member of the pair that is more likely to occur next") for 25 selected event pairs. Both estimation groups (PE and FE) plus a no-estimation control group (C) made the same series of choices; there were 12 subjects in each group.

Estimation performance was evaluated in terms of both error and correlation measures (relative to the "objective" values). Since the results were very comparable, only the error index is discussed here. As shown in Tables 6 the FE group was markedly

---

Table 6 about here

---

superior to the PE group. Both the group difference, $F(1,18) = 5.89$, $p < .03$, and the event type effects $F(7,126) = 2.99$, $p < .01$, were statistically significant. Performance on the subsequent choice task was considerably better for both estimation groups than for the control (see Table 7), and this effect, too, was

---

Table 7 about here

---

significant, $F(2,27) = 3.60$, $p < .05$.

2. The effect of prior cuing and estimation on choice. This study was designed to explore further the effect of estimation on choice. As noted earlier, the question was whether the facilitation is due chiefly to cuing or to summary encoding at retrieval. The

principal manipulation involved the crossing of frequency or probability <u>cuing</u> (via initial instructions) with frequency or probability <u>estimation</u> to form four estimation groups: FF, PP, FP, and PF. These groups were then compared to a no-estimation control (C) on the choice task.

Nine subjects were assigned randomly to each of the five groups. The resource-allocation problem was performed by all subjects in virtually the same form as in the previous study. The only changes, except for the prior cuing manipulation, were: (a) reduction of dispatching regions from 16 to 9, (b) reduction of the number of sessions from five to four, and of total calls from 600 to 225, (c) reduction of types of estimation probes from eight to five, and (d) increase in number of choice trials from 25 to 80, including 20 of the three-alternative variety. The distribution of calls is shown in Table 8, and the probe categories, in Table 9.

Tables 8 and 9 about here

Estimation performance for the four groups is shown in Table 10, and for the five probe types, in Table 11. Despite the pro-

---

Tables 10 and 11 about here

---

cedural changes, these data are remarkably similar to those obtained in the first study (see Table 6). Once again, FE was significantly better than PE performance, $F(1,32) = 71.03$, $p < .001$; and there was a highly significant event probe effect, $F(4,128) = 12.25$ $p < .001$. In this case, however, the interaction was also

significant, $F(4,128) = 26.89$, $p < .001$, a finding only _suggested_ by the previous data. Closer inspection of the items that contributed most to the FE-PE difference show them to be the more narrowly defined ones. The new variable in this study, congruity of cuing with estimation requirement, had no effect on estimation performance: $F(1,32) = 1.60$, $p < .210$.

The key predictive choice data are shown in Table 12. Here,

Table 12 about here

---

once again, it is clear that prior estimation aided performance, and by about the same amount as in the first study. Unlike estimation, however, the new congruity variable did appear to influence choice performance: accuracy under congruent cuing conditions averaged about 7% above that for incongruent conditions. While the overall difference among groups was highly significant, $F(4,40) = 4.77$, $p = .003$, the post hoc comparisons did not reveal significant differences among the congruent-incongruent pairs. Thus, while suggestive, the congruence effect must be interpreted with caution.

3. _Comparison of Direct Cuing With Prior Estimation Under Stationary and Nonstationary Conditions_. If the process through which estimation improves decision performance is one of general cuing, then one would expect the same effect with _direct_ cuing. Therefore, one variable in this study was the explicitness of cuing

(via instructions vs. estimation). A second variable was the consistency of the observed frequency evidence with prior generator beliefs. In the stationary condition, frequencies were consistent with descriptions of event "tendencies" presented at the outset of the experiment. In the nonstationary condition, the observed frequencies differed substantially from the prior descriptions (i.e. the "generator" was "shifted").

Cuing was manipulated between groups; stationarity, within groups. Thirteen subjects were assigned at random to each of the two groups. The direct cuing group was instructed, following the 225 dispatching trials, to apply the observed frequencies to the subsequent choice problems. The estimation group made frequency estimates (as in the previous studies) prior to the choice trials, but was not instructed to apply them to the choice task. Both groups observed both stationary and nonstationary events in the context of their dispatching exercise. In all other major respects, the format was identical to that used in previous studies.

The principal results are shown in Table 13. The group effect

---

Table 13 about here

---

did not approach significance, $F(1,24) < 1.00$, while the stationarity effect was highly significant, $F(1,24) = 27.64$, $p < .001$. The interaction was nonsignificant. Thus the predicted similarity of prior estimation to a direct cuing effect was established.

While the study was not designed to explore individual differences, it is of some interest to consider the accuracy of choices for individual subjects as a function of event stationarity and estimation performance.

Table 14 about here

As shown in Table 14, six subjects estimated the changed event frequencies about as well as the stationary ones. Six also performed well above chance on the choices involving changed events. Five of the six were common to both sets. Thus, as was discovered in the earlier studies using the personnel selection problem, there seem to be dramatic differences in individual propensities for processing frequentistic information; and those who process it are considerably more proficient in subsequent choice behavior.

III. Reports Generated Under the Project

1. Marques, T., & Howell, W. C. Intuitive frequency judgments as a function of prior expectations, observed evidence, and individual processing strategies. Technical Report No. 79-06, Department of Psychology Report Series, Houston: Rice University, 1979.

2. Marques, T. Attention control in the frequentistic processing of multidimensional event streams. Technical Report No. 80-13, Department of Psychology Report Series, Houston: Rice University, 1980.

3. Howell, W. C., & Kerkar, S.P. Uncertainty measurement in a

complex task as a function of response mode and event type characteristics. Technical Report No. 81-1, Department of Psychology Report Series, Houston: Rice University, 1981.

4. Kerkar, S.P., & Howell, W. C. Choosing among alternatives with uncertain outcomes: Effect of prior cuing and estimation requirements. Technical Report No. 81-2, Department of Psychology Report Series, Houston: Rice University, 1981.

5. Marques, T. E., & Howell, W. C. Processing frequency information under expanded task complexity: Limitations on automaticity. Under journal review.

6. Howell, W. C., & Kerkar, S. P. A test of task influences in uncertainty measurement. Under journal review.

REFERENCES

Einhorn, H. J., & Hogarth, R. M. Behavioral decision theory: Processes of judgment and choice. American Review of Psychology, 1981, 32, 53-88.

Fischhoff, B. Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, in press.

Hammond, K. R. The integration of research in judgment and decision theory. Center for Research on Judgment and Policy, Boulder, Colorado: Tech. Rep. No. 226, July, 1980.

Hammond, K. R., McClelland, G. H., & Mumpower, J. Human judgment and decision making: Theories, methods, and procedures. New York: Praeger, 1980.

Hasher, L., & Chromiak, W. The processing of frequency information: An automatic mechanism? Journal of Verbal Learning and Behavior, 1977, 16, 173-184.

Hasher, L., & Zacks, R. T. Automatic and effortful processing in memory. Journal of Experimental Psychology: General, 1979, 108, 356-388.

Howell, W. C., & Burnett, S. A. Uncertainty measurement: A cognitive taxonomy. Organizational Behavior and Human Performance, 1978, 22, 45-68.

Ross, L., & Anderson, C. A. Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In P. Slovic, A. Tversky, & D. Kahneman (Eds.). Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, in press.

Ross, L., & Lepper, M. R. The perseverence of beliefs: Empirical and normative considerations. In R. A. Shweder & D. Fiske (Eds.), New directions for methodology of behavioral sciences: Fallible judgment in behavioral research. San Francisco: Jossey-Bass, 1981.

Snyder, M., & Swann, W. B., Jr. Behavioral confirmation in social interaction: From social perception to social reality. Journal of Experimental Social Psychology, 1978, 14, 148-162.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. Annual Reviews of Psychology, 1977, 28, 1-39.

Tversky, A., & Kahneman, D. The framing of decisions and the psychology of choice. Science, 1981, 211, 453-458.

Table 1

Mean Unsigned Difference Between the Generated and Estimated Percentages of Target Events as a Function of Shift Magnitude and Experimental Sample Size.

Generator Shift	Experimental Sample Size			
(In percentage points)	20	40	60	80
0	10.15	5.50	8.75	9.25
10	12.60	9.75	12.25	7.45
20	17.25	20.05	13.50	14.75

Table 2

Frequency Estimation Performance

as a Function of Task Status.

Subject	Primary	Secondary
1	2.15	2.63
2	1.75	3.30
3	0.63	1.28
4	1.20	1.83

Table 3

Rating Performance ($r\ \hat{y} \cdot y$) as a Function

of Task Status.

Subject	Primary	Secondary
1	.896	.816
2	.952	-.060
3	.953	.891
4	.949	.904

TABLE 4

Distribution of Calls Over the 96 Event Categories Classified By Location, Type of Emergency and Level of Veracity.

	Type of Emergency					
	Police		Fire		Ambulance	
Location	AE	FA	AE	FA	AE	FA
1	2	1	-	-	2	-
2	1	-	4	1	-	-
3	1	-	-	-	-	-
4	-	-	-	-	-	-
5	10	3	-	-	4	1
6	8	2	-	-	9	2
7	7	2	-	-	4	-
8	10	5	3	-	6	-
9	2	2	2	-	7	3
10	-	-	7	1	10	1
11	10	2	-	-	1	-
12	3	4	1	-	-	-
13	-	-	-	-	-	-
14	-	-	1	1	-	-
15	-	-	-	-	-	-
16	-	-	4	-	-	-

Note: AE = actual emergency; FA = false alarm

Table 5

The Eight Categories of Information Probed With Illustrations of Comparable Items Administered to the Two Groups (FE and PE).

	Example*	
Probe Category	FE Group	PE Group
1. Type of event (Police, Fire, Amb.)	How many total police calls did you receive?	If a call comes in, what are the chances (0 - 100%) that it will be a police call?
2. Type by location.	(Map presented with instructions to estimate the totals for events indicated), e.g., police calls, zone 8.	(Map presented with instructions to estimate the chances, 0 - 100%, for events indicated), e.g., police call, zone 8.
3. Type by veracity.	How many false alarms were (police calls)?	Suppose a call was a false alarm. What are the chances of its being a (police) call?
4. Type by location by veracity.	(Map presented.) Please fill in the totals for false alarms only.	(Map presented.) Again, suppose a call was a false alarm. What are the chances it would be of the type and location indicated?
5. Response to veracity of calls.	On how many occasions did you verify a false alarm?	For any given call, what are the chances that you would verify a false alarm?
6. Response to event type.	How often did you correctly or incorrectly dispatch a (police) unit?	If you dispatched a unit, correctly or incorrectly, what are the chances that it was a (police) unit?
7. Correct response to event type by location.	(Map presented with instructions to estimate the number of units dispatched correctly for indicated type/location.)	(Map presented with instructions to estimate the chances that a unit correctly dispatched would be of indicated type/location.)
8. Correct response to event type.	How often did you correctly dispatch a (police) unit?	If you correctly dispatched a unit, what are the chances that it was a (police) unit?

*These examples are designed to give the reader a general idea of the probes used, not the exact format or total context of the questionnaire. For example, subject's understanding of the PE and FE response concepts.

Table 6

Mean Unsigned Error Scores for the Two Estimation Groups over the Eight Probe Categories.

	Group	
Probe Category	FE	PE
1. type of event	6.03	7.60
2. type by location	2.93	7.16
3. type by veracity	4.20	16.71
4. type by location by veracity	.98	7.17
5. response to veracity of calls	11.66	16.02
6. response to event type	9.93	12.61
7. correct response to event type by location	1.45	18.83
8. correct response to event type	6.38	12.04

Table 7

Predictive Choice Accuracy and Overall

Estimation Performance for the Three

Groups.

	Percentage	Average Estimates	
Group	Correct Choices	Correlation	Unsigned Error
FE	77.6	.75	5.45
PE	71.6	.54	12.27
Control	59.2	--	---

Table 8

Distribution of Emergency Calls Over 36 Event Categories Classified By Location, Type of Emergency and Level of Veracity.

	Type of Emergency			
	Police		Fire	
Location	AE	FA	AE	FA
1	4	4	1	1
2	2	0	1	0
3	0	0	0	2
4	2	2	10	0
5	8	2	1	0
6	0	1	8	0
7	4	4	0	0
8	8	2	0	0
9	0	4	2	2

Note: AE = actual emergency

FA = false alarm

Table 9

Illustrations of Items in the Five Categories of Information Probed for the Two Estimation Tasks (Frequency and Probability).

	Example	
Probe Category	Frequency Estimation (FF and PF groups)	Probability Estimation (FP and PP groups)
1. Type of Event	How many total police calls did you receive?	If a call comes in, what are the chances (0-100%) that it will be a police call?
2. Type of Veracity	How many total false alarms did you receive?	If a call comes in, what are the chances that it will be a false alarm?
3. Event Type by Veracity	How many false alarms were police calls?	Suppose a call was a police call. What are the chances of its being a false alarm?
4. Event type by Location	(Map presented with instructions to estimate totals for events indicated), e.g., police calls, sector 1.	(Map presented with instructions to estimate the chances, 0-100%, for events indicated), e.g. police call, sector 1.
5. Event Type by Location by Veracity	(Map presented). Please fill in totals for false alarms only for events indicated, e.g. fire call, sector 2.	(Map presented). Suppose a call was a fire call. What are the chances it would be a false alarm in the indicated location, e.g. sector 2?

Table 10

Mean Error Scores for the Four Experimental Groups on the Deviation Measures Across all Probe Categories.

Group	Unsigned Error
FF	4.28
FP	16.68
PF	4.18
PP	13.51

Table 11

Frequency and Probability Estimation Performance on the Five Probe Items With Respect to the Deviation Measures.

Probe Category	Frequency* Estimation	Probability** Estimation
	Unsigned Error	
1. Type of Event	6.28	6.39
2. Type of of Veracity	7.06	9.61
3. Event Type by Veracity	4.61	11.76
4. Event Type by Location	2.12	14.12
5. Event Type by Location by Veracity	1.07	33.59

*Frequency Estimation refers to the performance of FF and PF groups.

**Probability Estimation refers to the performance of FP and PP groups.

Table 12

Mean Predictive Choice Accuracy and Overall Estimation Performance for the Five Groups.

	Percentage Correct Choices*		Estimation
Group	2-Choice	3-Choice	Unsigned Error
FF	77.04	70.56	4.28
FP	71.30	65.00	16.68
PF	72.78	61.67	4.18
PP	75.00	74.44	13.51
C	61.67	45.00	---

*Note: These are raw accurate scores. One must correct for chance in order to make meaningful comparisons between the 2- and 3-choice cases. All analyses used corrected scores, a procedure that rendered the apparent 2 vs. 3-choice difference trivial and nonsignificant.

Table 13

Average Predictive Choice Accuracy for the Four Experimental Conditions (Raw Scores).

Group	Event Generator: Stationary	Event Generator: Non-Stationary
Estimation	76.63	55.62
Direct Cuing	75.15	62.43

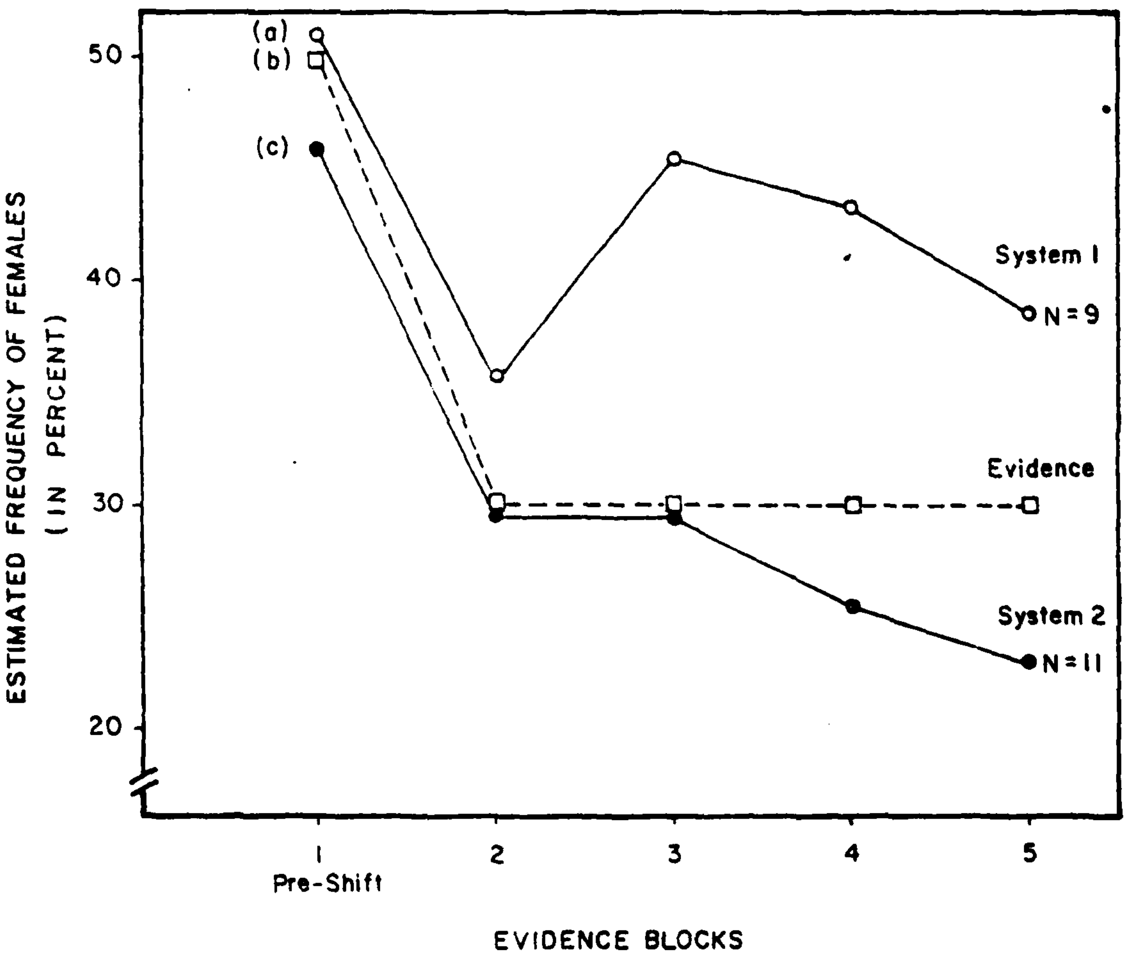
Table 14

Individual Performance on Estimation and Choice Under Stationary and Non-stationary Generator Conditions (Estimation Group Only).

	Estimation				Choice	
	Error		Correlation		Accuracy	
Subject	Stat.	Non-Stat.	Stat.	Non-Stat.	Stat.	Non-Stat.
1	1.70	1.38	.76	.93+	84.62	88.46*
2	1.90	2.75	.78	.39	76.92	42.31
3	2.00	1.75	.74	.75+	61.54	57.69*
4	1.60	2.00	.85	.71+	76.92	38.46
5	2.30	2.63	.75	.24	73.08	46.15
6	1.90	3.13	.91	-.17	76.92	50.00
7	1.70	3.38	.74	-.33	69.23	65.38*
8	1.80	3.63	.86	-.12	80.77	46.15
9	2.80	3.00	.68	.78+	80.77	57.69*
10	2.70	1.75	.67	.82+	65.38	57.69*
11	1.80	2.50	.85	.50	88.46	53.85
12	1.60	1.13	.78	.88+	73.08	76.92*
13	.70	2.63	.96	.05	88.46	42.31

+Individuals who responded well to the generator shift in their estimation performance.

*Individuals who made above-chance accuracy on predictive choices for the shifted generator.


ESTIMATED FREQUENCY OF FEMALES (IN PERCENT)
50
40
30
20
(a)
(b)
(c)
System I
N = 9
Evidence
System 2
N = 11
1
Pre-Shift
2
3
4
5
EVIDENCE BLOCKS